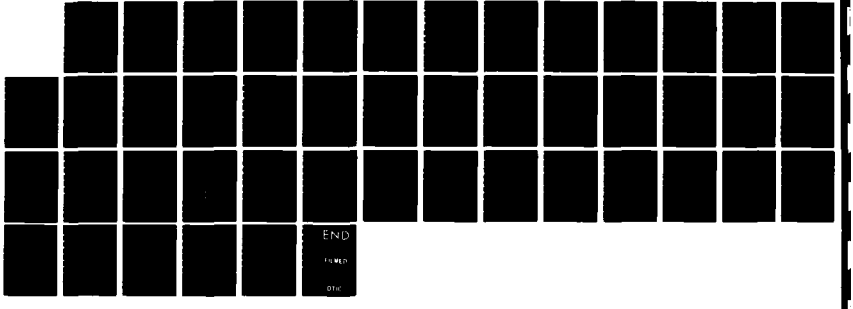
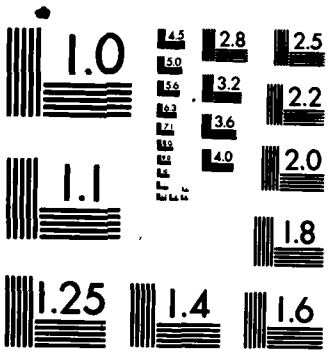
AD-A161 874 THE GEONAMES PROCESSING SYSTEM FUNCTIONAL DESIGN SPECIFICATION VOLUME 4 A (U) NAVAL OCEAN RESEARCH AND DEVELOPMENT ACTIVITY NSTL STATION MS G LANGRAN 1/1

UNCLASSIFIED MAR 85 NORDA-101 F/G 9/2 NL





(12)

Naval Ocean Research and Development Activity
NSTL, Mississippi 39529 NORDA Report 101 March 1985

Geonames Processing System Functional Design Specification

Volume 4: Advanced Symbol Processing

Final Report

Gail Langran

Mapping, Charting, and Geodesy Division
Ocean Science Directorate

FOREWORD

DMA has recognized a need for digital procedures to store, retrieve, and edit geographic names and to prepare names data for product generation. DMA's stated goal is a 50-100 million name digital data base with subsystems to capture names data, edit and format data, and prepare names overlays for maps. NORDA began a geonames processing system design study late in FY82. This report is one of a five-volume series of reports that describes the functional design of the digital geographic names processing system.

R. P. Onorati, Captain, USN
Commanding Officer, NORDA

This NORDA Report was prepared to meet style requirements of the sponsor.

EXECUTIVE SUMMARY

In FY82, the Pattern Analysis Branch, Mapping, Charting and Geodesy Division of the Naval Ocean Research and Development Activity (NORDA) began work for the Defense Mapping Agency (DMA) on four interrelated aspects of computer-assisted geographic names processing:

- digital capture of names and named feature information from analog sources such as maps and gazetteers;
- adaptation of a data base management system for a very large, product-independent set of world geographic names and their descriptors to support a variety of DMA products and applications;
- word and symbol processing to include editing text with diacritics and special symbols, and document formatting;
- digital type layout on maps, gazetteers, and other DMA products with associated data selection, formatting, scaling, and type generation.

DMA's four original requirement statements are in Appendix B.

This work, referred to as the Geonames Processing System, will be conducted during FY82-FY89. A Comprehensive Coordination Plan (NORDA Technical Note 189) was written in FY82 to provide a general system description. The Geonames Processing System Functional Design Specification, of which this is Volume 4, describes in more detail a proposed system based on requirements presented to NORDA by DMA Headquarters and DMA's two production centers.

Concurrent and related development by DMA's Special Program Office for Exploitation and Modernization is not considered in this set of reports. Rather, a comprehensive system with simple and adaptable standard interfaces is described.

This report states the functional specifications of Advanced Symbol Processing, a specialized word processor for text containing diacritics and special symbols, hardware-connected to a geographic names data base.

Section 1 is an overview of the word and symbol processing subsystem. Section 7 shows its overall processing flow. Sections 2-6 discuss subsystem functions. Interfaces between the four geonames processing subsystems are in Section 9, and permanent data sets are in Section 8. Performance and hardware specifications are in Sections 11 and 12, respectively.

ACKNOWLEDGMENTS

This work was sponsored by DMA under Program Element 64701B, with subtask title, "Geonames Processing System." Mr. Dennis Franklin and Lt. Col. Tom Baybrook, both of DMAHQ/STT, shared project management duties during the writing of this report. Their help in communicating with DMA's production centers and providing information on DMA production methods was instrumental to this functional design. Dr. Don Durham, head of NORDA's Mapping, Charting, and Geodesy (MC&G) Division, and Dr. Charles Walker, head of the MC&G Division's Pattern Analysis Branch, contributed valuable advice and assistance. Ms. Anne Downs, Dr. Berry Glick, and Mr. Warren Schmidt, coauthors of Volume 1 (Automated Alphanumeric Data Entry System) and Dr. Allen Barnes and Mr. Steven Miller, coauthors of Volume 2 (Geographic Names Data Base) deserve thanks and credit for discussions and ideas that led to design and definition of the data entities, data sets, and interfaces shared among the four subsystems.

TABLE OF CONTENTS

FIGURES

TABLES

INTRODUCTION

a. Organizations

Defense Mapping Agency Headquarters (DMAHQ)
U.S. Naval Observatory
Washington, D.C.

Defense Mapping Agency Hydrographic/Topographic Center (DMAHTC)
6500 Brookes Lane
Washington, D.C.

Defense Mapping Agency Aerospace Center (DMAAC)
3200 South Second St.
St. Louis, Missouri

b. Scope

The purpose of this report is to describe system attributes, serving as a basis for mutual understanding between the user and the developer.

The Geonames Processing Subsystems are often referred to in this report by their acronyms: ASP (Advanced Symbol Processing); ATP (Advanced Type Placement); GNDB (Geographic Names Data Base); and AADES (Automated Alphanumeric Data Entry System).

c. Background

In FY82 the Pattern Analysis Branch, Mapping, Charting, and Geodesy Division of the Naval Ocean Research and Development Activity (NORDA) began a subtask for the Defense Mapping Agency (DMA) entitled "Advanced Type Placement and Geonames Data Base System Development," a project encompassing the digital capture, storage, edit, and display of geographic names. The subtask in its current form is an amalgamation of four previous DMA requirements for independent development of a geographic names data base, a system for high-volume geographic names data capture, advanced word and symbol processing, and automated type placement for maps (see Appendix B for DMA's original requirement statements). A Comprehensive Coordination Plan was submitted by NORDA as a preliminary definition of the overall Geonames Processing System subtasks and their interfaces.

d. Description

The complete Geonames Processing System is comprised of four components (Fig. i-1).

- The Automated Alphanumeric Data Entry System provides a means of high-volume geographic names data capture. World geonames with their corresponding locations and attributes will be captured from both tabular and map/chart sources using raster scan and optical character reading technologies. AADES converts alphanumeric data into computer-readable form with a 99% accuracy rate. It requires minimum operator intervention, pro-

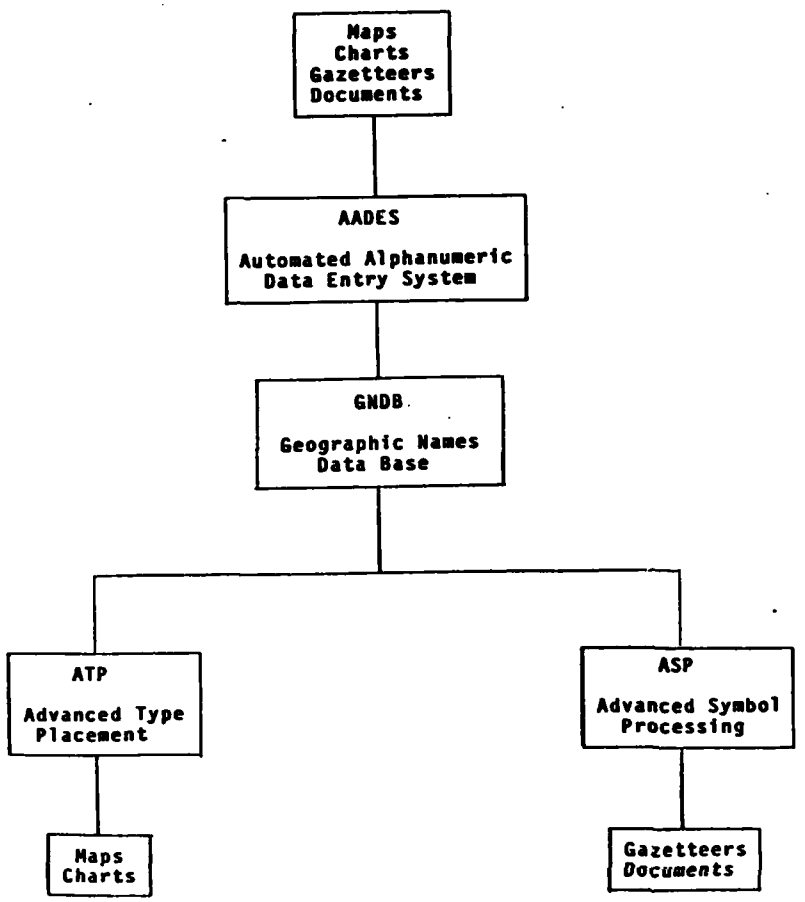


Figure i-1. Geonames Processing System overview

vides automated error checking, and results in clean data files for supervised merging with the GNDB.

- Geographic Names Data Base stores world geonames and their descriptors in non-product-oriented files. It provides extensive query capabilities to support data base updates, chart and gazetteer compilation, and toponymic research. The GNDB's ultimate size will be 50–100 million geonames.

- Advanced Symbol Processing. International geonames comprised of diacritics and special symbols require specialized hardware and software for access, manipulation, and editing. ASP provides alphanumeric edit and display of world geonames and advanced word processing capabilities such as sorting, searching, and formatting.

- Advanced Type Placement automates the production of map names overlays by exploiting electronic display technology and the rule-based nature of cartographic names placement. ATP includes automated utilities for names selection, type composition, type placement, virtual map display, and interactive graphic edit.

The Geonames Processing System responds to a major need: it will integrate DMA's names processing tasks into the digital map production pipeline and coordinate all geonames processing activities. Obvious benefits are increased production rates and lowered costs. Overall accuracy and coverage

should also improve with the increased efficiency of such a system. Helpful utilities will raise toponymic researchers' productivity levels. Further automation will be easier to implement once the process is converted from manual to digital.

This Functional Design Specification addresses the development of Advanced Symbol Processing for international geonames (ASP). Overall processing flow is outlined in Section 8. It is suggested that the reader turn to this section before continuing.

e. Applicable Documents

The following references provide a summary of the basis for the Geonames Processing Subtask development.

- "Advanced Type Placement and Geonames Database: Comprehensive Coordination Plan." NORDA Technical Note 189, January 1983.
- "A Prototype Geographic Names Input Station for the Defense Mapping Agency." Paper presented at Auto Carto IV by D.R. Caldwell and D.E. Strife, September 1982.
- "Names Type File System." Consulting Report for the U.S.A.E.T.L Project #POO13, April 1983.
- "Development of an Automated Cartographic Capability." The Final Report of the Automated Cartography Task Force, Defense Mapping Agency Hydrographic/Topographic Center, April 1982.
- "The Feasibility of Establishing an Automated Chart Production Process." The Defense Mapping Agency Hydrographic/Topographic Center, October 1982.

f. Limitations

The individual subsystem descriptions are functional and not physical definitions, i.e.:

- a given function required by a given subsystem may not be performed upon the hardware logically associated with the subsystem,
- one software module may serve several of the subsystem functional requirements.

Thus the functions and data sets specified in this five-volume set of design specifications are described somewhat redundantly to fully define each subsystem regardless of its interplay with other subsystems. Physical (hardware and software) synthesis will be accomplished and described by the Implementation Plan at a later date.

1.0 ADVANCED SYMBOL PROCESSING OVERVIEW

Advanced Symbol Processing (ASP) is for text editing, interactive read/write data base access, low-volume data entry, data formatting, and proof printing of geonames data. ASP supports data base maintenance and geonames product generation (mainly compiling and publishing gazetteers and other tabular documents).

ASP will:

- process text containing diacritics and special symbols,
- provide utilities to aid in toponymic analysis, and
- interface to a computer typesetter for final printing.

1.1 Existing Methods and Procedures

The Names Input Station (NIS) is a prototype for the Advanced Symbol Processor. The NIS is comprised of a Plessey PDP-11/70 with a 2-MB disk connected to an ECD intelligent terminal. The NIS uses two outboard keypads for single- or dual-keystroke entry of diacritics and special symbols. Regional Diacritics Sets (REDS) partition diacritics by country linguistic requirements.

1.2 Deficiencies

The NIS has serious hardware limitations. Down-time is high. ECD has periodically varied its intelligent terminal configuration, necessitating software adjustments each expansion. Because the ECD processes text only by screenful text manipulation, involving diacritics is limited to approximately 122 lines. Thus, global changes and sorting are arduous.

NIS operating procedures and keyboard layout are adequate, but could be improved by devices to help operators remember function key layout. The REDS strategy is successful. REDS, however, do not fully meet DMA's areal requirements and must be expanded.

Portions of NIS software are coded in a macro language that is difficult to maintain in-house or by contract. The Multiset III/NIS interface is poor. Although NIS tapes can be read by Multiset III, Multiset III tapes cannot be read by the NIS.

1.3 Proposed Methods and Procedures

A multiple-station system tied to the Multiset III computer typesetter is required. ASP will be a minicomputer processor accessed by graphics terminals capable of diacritics text processing and read/write GNDB access. A modem link from ASP to Multiset III will be investigated.

Software will be written in a high-level programming language, incorporating the NIS software (mostly FORTRAN IV coded) whenever possible. Output formats of the current system will be maintained, as will be the Regional Diacritics Sets. Improved REDS codes developed by DMA will be incorporated.

1.4 Standard Terms

The following is a key to ASP Terminology.

- Regional Diacritics Sets (REDS): 14 sets of diacritics are grouped according to the linguistic needs of various geographical subdivisions.

- Names Input Station (NIS): a prototype diacritics text processor currently used by DMA in support of gazetteer production.

- Names Data File: a file of specifiable format that holds names data in the process of revision.

- Standard Data Transfer Record: a standard interface format used by the four geonames processing subsystems.

- Gazetteer Tape: a digital gazetteer, routed to Multiset III for typesetting and printing.

- Foreign Place Names File (FPNF): a card file maintained by DMA containing approximately 4.5 million geonames and associated feature data.

- Feature Designator: a three-to four-letter code describing a feature type (i.e., PPL means populated place), currently used in DMA gazetteers.

- Feature Attribute: non-nominal feature data, i.e., population, stream discharge, area, perimeter, or importance.

- Map/Chart: are used interchangeably.

- Data Base Update File: potential GNDB revisions to be verified by a toponymist and written to the data base.

1.5 Ordering of Functional Design Specification

Section 7 provides a detailed description of ASP data flow. ASP performs five major functions:

- specialized word processing,
- GNDB read/write access,
- output processing,
- job management, and
- file management.

Word Processing includes formatting and sorting, toponymic utilities, and basic text editing functions for text with diacritics and other foreign symbols. The GNDB is accessed from an ASP work-station due to ASP's diacritics-handling abilities. ASP alone has read/write GNDB access for concurrency control and quick interactive response. ASP allows typed entry of new names data when low volume does not justify using AADES. Output Processing is an administrative function initiated by the analyst upon completion of a production job. Job Management controls the ASP work flow and user interface. File Management oversees file storage and retrieval, and catalogs ASP digital product archives.

ASP functions are described in Sections 2-6. Section 8 describes ASP data sets. Section 9 defines the interfaces between ASP and the other three Geonames Processing subsystems. Section 10 lists assumptions and constraints concerning system design. Sections 11 and 12 specify performance and hardware requirements, respectively. Performance specifications shared by the four Geonames Processing Subsystems are described in Volume 5 of this series. These specifications include software requirements, human engineering requirements, documentation requirements, training, acceptance testing, system maintainability, and miscellaneous requirements.

2.0 WORD PROCESSING

Word Processing edits text with diacritics and special symbols. It sorts, merges, and formats lists. It provides utilities to aid in toponymic research and names data manipulation.

2.1 Text Editor

The text editor will add, delete, tabulate, search for a string, make global changes, and copy or move blocks of text to other portions of the document in progress. The analyst initiates these operations with well-defined function keys or mnemonics.

2.2 Formatter

The formatter modifies fields of tabular data to organize documents. It can sort tabular data by any field either alphabetically or numerically, as commanded. Data fields are displayable in any columnar order.

For convenience, commonly used formats are storable. A default format (i.e. the format used by the system when no format was specified) can be set.

The formatter merges stored files with files in memory. It provides utilities that manipulate data, described below.

2.2.1 Coordinate Transformation

Coordinates can be converted from UTM to geographic coordinates and vice versa. Latitudes and longitudes can be rounded to the nearest minute, and decimal values can be converted to degrees, minutes, and seconds.

2.2.2 Map Sheet Lookup

Given a geographic location and a map series, this function finds the sheet number of the map covering the specified location.

2.2.3 Metric Conversion

This function converts statute miles, nautical miles, yards, feet, inches, and acres to a specified metric unit and vice versa.

2.3 Toponymic Aids

Transliteration help sequences use algorithms and lookup tables to provide on-line assistance to toponymic researchers, as described below.

2.3.1 Telegrapher's Code

The Telegrapher's Code is an on-line reference table that for reversing the Romanization of ideograms.

2.3.2 Transliteration Rules

The transliteration rules shown in gazetteer forewords are kept on-line for reference. Transliteration rules are interactively consulted or added to Gazetteer Files.

2.3.3 Dictionary

The Dictionary File defines all Feature Designators and all generics used in geonames processing. It handles the following.

- What is the definition of (term)?
- Which feature designations are (hydrographic, hypsographic, cultural)?

2.4 Input

The text editor operates on files in memory. Other file usage may involve:

- Coordinate Limits File,
- Map Sheet Boundaries File,
- Telegrapher's Code File,
- Transliteration Rules File, or
- Dictionary File

2.5 Output

Output is saved as a Names Data File, a Map Data File, a Gazetteer File, or a Data Base Update File. Format and field information, date, and analyst name are written to the header.

3.0 DATA BASE MANIPULATION

ASP is the only Geonames Processing subsystem with read/write data base access. Data Base Manipulation controls:

- read-only GNDB queries and Names Data File compilation,
- read/write Data Base Update File merges.

Data Base Manipulation contains four major subfunctions (described in more detail below):

- data base querying,
- file compilation,
- data base updates and maintenance, and
- product statistics.

3.1 Data Base Queries

Low-volume queries and geonames research use an interface that facilitates query formulation. GNDB queries precede file compilation to ensure correct file contents (Fig. 3-1), and are made during GNDB updates.

A query takes the following form: target data are specified by data fields and parameters. Logical concepts to be supported are described in this section. A "report" option returns the number of qualifying entries ("hits") to avoid compiling incorrectly targeted data. If the number of hits corresponds to the analyst's expectations, he/she specifies a destination, i.e. terminal or filename, and the qualifying data are written to that location. If the number seems unreasonable, the analyst narrows or attempts to correct the targeting instructions.

3.1.1 Target Data Specification

Target data are logically delimited by search parameters and data fields. Field specification may entail a single field (perhaps only the name of a place or a maximum population) or many fields, as required.

Search ranges are delimited logically and may restrict by feature designation, area, or spelling. Queries may be expressed using tables or mnemonic commands. The system handles general concepts, i.e., all cultural placenames, all hydrographic placenames, or all hypsographic placenames. Some examples of logical expressions are:

- All names, accompanied by coordinates, of populated places larger than 50,000 in Canada;
- All names, accompanied by feature designator and coordinates, of hydrological features in England.

Expressions that target by area and by spelling follow.

3.1.1.a Locate Names in an Area

A target search area may be specified in relative or in absolute terms. A relative specification is a radial distance:

- within 20 miles of London, or
- within 50 miles of the mouth of the Columbia River.

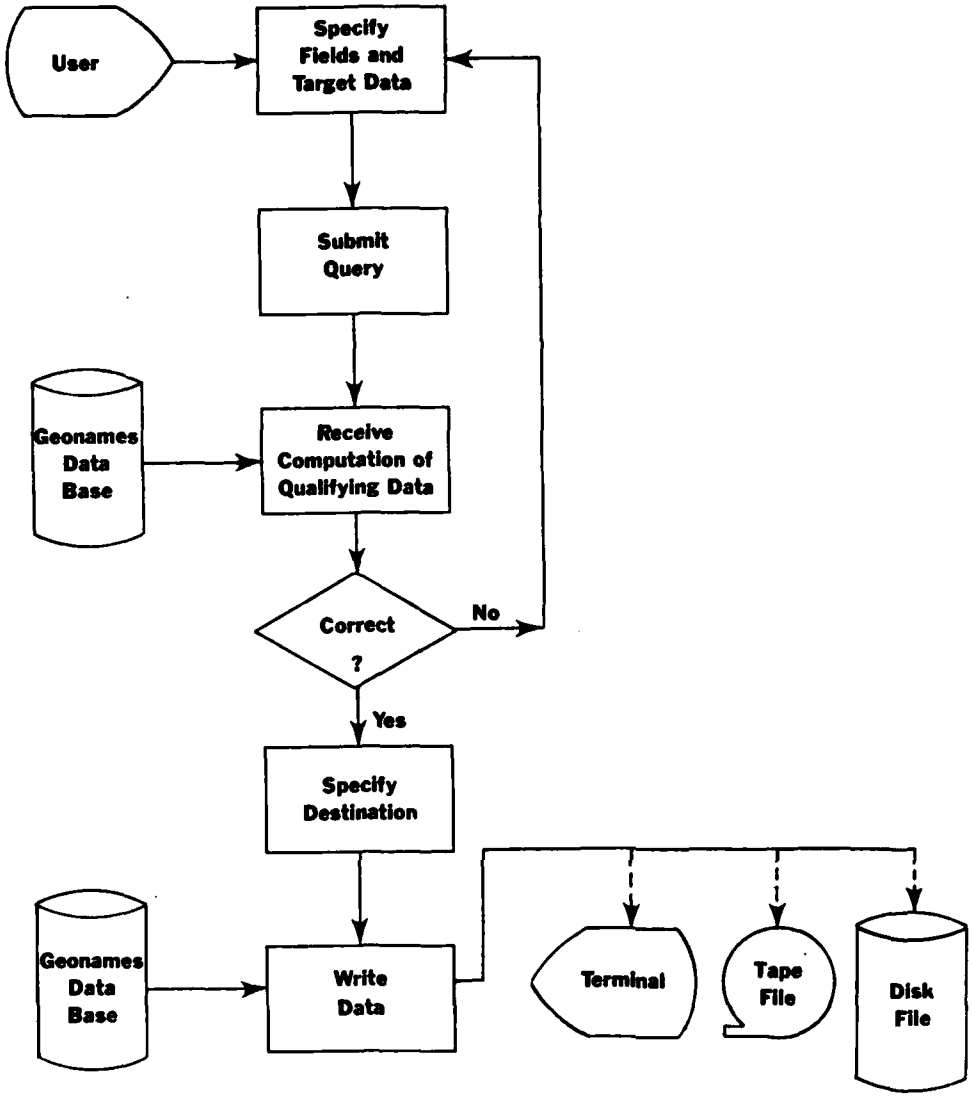


Figure 3-1. Data Base Query

An absolute location is a coordinate window, a map sheet, or a geographic region. A coordinate window may be specified by lower left and upper right corners, or by lower left corner and length of sides. Some examples of absolute area specifications:

- in a window bounded by (X_1,Y_1), (X_2,Y_2),
- in Turkey,
- in a 100 X 100 mile window with lower left (X_1,Y_1), or
- on (map code, map sheet).

3.1.1.b Locate Names of Similar Spelling

To further narrow a data set, a string-matching utility matches characters of targeted geonames to a user-supplied string comprised of alphanumerics and "wildcards." Target data are specified as described above.

A wildcard represents any number of unspecified characters.

Gib*	matches	Gibara Gibeon Gibraleon Gibraltar Gibson
Gib*n	matches	Gibeon Gibraleon Gibson
Gib*a*n	matches	Gibraleon

A limited wildcard represents one unspecified character.

Gib???	matches	Gibara Gibeon Gibson
Gib?on	matches	Gibeon Gibson
Gib????on	matches	Gibraleon

3.2 Compilation From the GNDB

Data sets to be compiled are targeted the same way as for querying, described in Section 3.1. Both interactive and batch file compilation are possible.

Files are compiled interactively to the terminal or to a specified data file by querying the data base until the correct target data set is assured. Then, compilation is approved and initiated. A "notify" option alerts the terminal when compilation is completed. Query techniques described above also target data for batch compilation. Data are targeted interactively, then queued for processing. Or, a logical expression is coded in the batch job.

3.3 Data Base Updates

Data Base Update Files are merged with the GNDB interactively. New data are checked by toponymists to assure that they are correct and unique. Incoming data are checked against stored data because information sources may conflict due to regional differences, publication date, or scale. Some automation, described next, is provided to assist toponymists.

3.3.1 Structural Constraints

Structural constraints protect data base integrity. Duplicate records and duplicate paths are disallowed.

3.3.2 Sanity Checks

A sanity check tests data values for validity. Data that fail the test are considered to be in error. Safety measures to prevent bad data from GNDB entry will include a pre-execution echo to

provide the analyst a final look at each data entry before admission to the GNDB. The following sanity checks help avoid data entry errors.

3.3.2.a Reasonable Bounds

Potential GNDB entries are checked against stored Reasonable Bounds, changeable only by the data base manager. Some examples:

- settlement population: 1-10 million;
- elevation: 0-29,100 feet;
- length of rivers: 1-4200 miles;
- latitude and longitude limits: 0 to 90 North or South (stored as −90 to 90) and 0 to 180 East or West (stored as 0 to 360);
- city boundary size: less than 50 miles.

3.3.2.b Relational Integrity

Known relationships are checked by the software. At minimum, the software tests that:

- the position of an entity is within its country;
- a stated or implicit inclusion relation does, in fact, exist.

3.3.2.c Consistency Assertions

Consistency prohibits storing mixed attribute measurements. If one feature's length is described in meters and another's in centimeters, a misleading query response results. A measurement, when displayed by the system, must be followed by its unit indicator to bring the system's assumption to the analyst's attention.

3.3.3 Alias Detection

Before adding to the GNDB, new names are examined to assure they are original and not variant names or spellings of previously captured data. Utilities help toponymists with this task.

3.3.3.a Toponymic Aids

The on-line toponymic help sequences shown below are supplied.

- Telegrapher's Code—a reference table that allows reversal of Romanization from ideograms.
- Transliteration Rules—transliteration rules as shown in gazetteer forwards. Transliteration rules are displayed on request at the terminal.
- Dictionary—defines feature designators and generics used in processing. It can be queried in several ways:
  - what is the definition of (term)?
  - which feature designators are (hydrographic, cultural, hypsographic)?

3.3.3.b Searching a Geographic Area

The query utility that locates names within an area (Section 3.1.1.a) also limits the search area for duplicate names.

3.3.3.c Searching for Variant Spellings

The query utility that locates names with specified spellings (Section 3.1.1.b) will also search a specified file for duplicate spellings, using syntax and protocol described earlier.

3.3.3.d Coordinate Transformation

This utility converts coordinates from UTM to geographic coordinates, and vice versa. It also rounds latitudes and longitudes to the nearest minute, and converts coordinates from decimal to degree, minute, second values.

3.3.3.e Metric Conversion

Mathematical functions for converting statute miles, nautical miles, yards, feet, inches, and acres to the metric unit specified and vice versa are provided.

3.4 Product Statistics

Each GNDB revision outdates one or more of DMA's products. Statistics are kept on the number of name changes in each product's area. When a new name is entered or a name changes, the affected products are located using a map sheet lines file or country name. Then, name changes are tallied in the Product Statistics Files of all affected products.

3.5 Input

The GNDB is input to all Data Base Manipulation. The Data Base Update File is input to all GNDB updates. Internal files used according to the functions selected by the analyst are:

- Telegrapher's Code File,
- Dictionary File,
- Transliteration Rules File,
- Map Sheet Boundary File,
- Product Statistics File,
- Coordinate Limits File.

3.6 Output

Data Base Query output goes to the terminal or a file. Output from Data Base Updates are GNDB revisions and entries to Product Statistics Files. Output from Batch File Compilation is a Names Data File with header describing file format.

4.0 OUTPUT PROCESSING

The analyst initiates output processing upon completion of a product. Output processing is batch; its operations follow.

- A Gazetteer File (or other product-specific file) is created for the completed product. The previous edition's Gazetteer File is optionally deleted.
- The new file's header information is written to the Archives Catalog.
- Completion notices are written to the System Log File and the Product Log File.
- Upon verification of the new file created in step one, the Names Data File used for its generation is destroyed.
- The product is scheduled for hardcopy reproduction.

4.1 Input

Input is a Names Data File.

4.2 Output

Primary output is the new Gazetteer File or product-specific file. Entries are written to the System Log File, the Archives Catalog, and the Product Log File.

5.0 JOB MANAGEMENT

Job Management controls ASP processing flow. It forms an interface between the user and the various system functions. It also maintains records on work scheduled, work in process, and work completed.

Job Management records the time and date that each function was performed on a given data set. Interactive records include the editor and the edit system. Batch records include the initiator. All functions records include the start time and stop time. These records describe the processing stage of all data sets to the production supervisor.

5.1 Product Log

For each Names Data File in process the following records are maintained in the Product Log File:

- processing history to include the following information for each job performed on the data set:
  - production supervisor,
  - editor and edit station,
  - time and date,
  - contributing data sets,
  - parameters, if applicable, and
  - purpose;
- the next processing step required;
- a message file.

5.2 System Log

A System Log, maintained by Job Management, provides daily, weekly, and monthly statistics, including:

- process initiation entry,
- process termination entry,
- tape initiation entry,
- tape termination entry,
- disk initiation entry,
- disk termination entry,
- edit station initiation entry,
- edit station termination entry,
- system supervisor console entry, and
- system utilization.

The System Log File is described in more detail in Section 8.

5.3 Input

The production supervisor inputs through his console all data, commands, and inquiries, including:

- job step sequences and priorities,
- job assignments to resources,
- status report requests, and
- inquiries.

5.4 Output

Output goes to the Product Log File and the System Log File. Output includes:

- listing of work in progress (audit trail),
- system resource commitments, and
- responses to queries.

6.0 FILE MANAGEMENT

6.1 File Management Functions

File Management stores and retrieves ASP data sets. Every ASP task requiring access to mass storage has access to the file manager, whose functions follow.

- Manage the data flow to and from disk and overflow tape storage.
- Allocate disk file space.
- Act as the interface between central mass storage and all processing functions.
- Enforce security and other access protection requirements.
- Record status and location of all data sets, whether on disk, in temporary overflow storage, or archive tape.
- Create, delete, and update data sets.
- Back up all data.
- Answer requests for data set status and location information from the production supervisors and operators.
- Retrieve specific information on all data sets within a geographic area defined by a rectangle.
- Transfer a data set to an archive tape when so requested by Job Management.
- Permit retrieval of data set by its identifiers, which include data set ID, map sheet, and project ID.
- Transfer data sets between workstations and Output Processing when the data set is not stored first.

6.2 Internal Data Sets

ASP uses a variety of internal data sets. File Management maintains these data sets until the supervisor gives explicit instructions for their disposition via Job Management. Certain standard information is recorded for each type of data set:

- internally assigned data set identifier,
- security classification, and
- applications.

7.0 PROCESSING FLOW

Advanced Symbol Processing takes different forms based on the task at hand. Input is from:

- the GNDB,
- an archive tape,
- a Data Base Update File,
- the workstation keyboard, or
- a Names Data File.

Names Data Files are gazetteers, names overlays, or toponymic investigations in progress.

Important ASP functions are word processing and data base access. The ASP application dictates processing flow. File preparation, queries, and data base updates require separate processing paths, described here.

7.1 File Preparation

File preparation uses ASP word processing functions to revise or reformat a file, using the following steps.

7.1.1 Get a Workfile

Workfiles are compiled or generated in the following ways:

- compilation from GNDB (Section 3.4),
- compilation from tape,
- typed entry from the ASP workstation (Section 2.1),
- recall from disk using file manager.

7.1.2 Reconcile Conflicting Data

If two or more files were merged in the previous step, duplicate records are located and purged. Names with spelling, locational, or other discrepancies are interactively sought and modified as described in Section 3.3.1.

7.1.3 Edit and Format

The workfile is examined, corrected, and otherwise manipulated. Editing prepares it for the next processing step, which can be:

- computer typesetting for tabular geonames product,
- names overlay production, or
- further query and research.

7.2 Interactive Queries

ASP is not always product-oriented, and its data sets are not always retained. Interactive queries do not require a workfile unless records are needed. Data base queries are parameterized by logical expressions. Responses to queries are routed to the terminal, a file, or a hardcopy device, as specified.

7.3 Data Base Updates

ASP is the window into the GNDB. As such, all GNDB additions or corrections are made by a toponymist using ASP and GNDB utilities. Data Base Update Files are generated by AADES, ATP, and ASP, and are input to the GNDB using ASP's Data Base Update function. When conflicts exist or data correctness is in doubt, ASP utilities help a toponymist to find errors. Each entry in the Data Base Update File is described by source, originator, and previous research progress.

A Data Base Update File originating from AADES contains data from a single source and area. AADES Data Base Update Files are unresearched. The toponymist might begin by compiling subfiles, perhaps of city names or areal feature names. Then, he/she queries the subfile for similar spellings or similar features within a given radial distance. The system responds with the number of "hits" after each query to allow the option of immediate display or further target refinement. Toponymic aids relieve the toponymist of repeated searches through a data list. Decisions on data correctness, however, are left to the toponymist.

A Data Base Update File originating from ATP is usually fully researched new names of a single geographic area. Merging this file with the GNDB requires only nominal supervision. Since ATP begins with a file compiled from the GNDB, names in an ATP Data Base Update File are presumably reconciled with GNDB names. The data base is searched for the field to be replaced, the full name record is displayed, and the analyst prompted for approval.

A Data Base Update File originating from ASP may be names from many areas and different sources, in different stages of research, entered through the ASP workstation keyboard. Such a file may be in a constant state of change as each name is corrected, entered into the GNDB, and purged from the Data Base Update File.

8.0 DATA SETS

8.1 Processing Data Sets

Processing data sets are created by ASP functions, then used or modified by ASP. None are retained upon completion of processing.

8.1.1 Names Data File

The Names Data File is a toponymic workfile compiled from the GNDB at an ASP workstation. It is the sole read interface between ATP and the GNDB, and the sole batch read interface between ASP and the GNDB.

8.1.1.a Header. The header denotes file contents and format (both are specified when the file is created). Required are:

- data fields and format,
- compilation date,
- purpose of file,
- date of each use and the analyst involved, and
- comments.

8.1.1.b Contents. Names Data Files compiled from the GNDB include a set of requested geonames data for a given area. This can include one or all of the entities contained in the GNDB (see Volume 2, Section 5.1). Files are likely to include:

- named feature coordinates,*
- placename string,
- feature designator,
- positional resolution,
- feature attributes,
- variant spellings,
- reference source and date, and
- transliteration code.

8.1.2 Data Base Update File

Data Base Update Files are formatted to expedite routine toponymic comparisons and merging of same-area data sets gathered from different sources, including those compiled from the GNDB. When discrepancies are reconciled and new information is added, the files undergo supervised entry to the GNDB.

8.1.2.a Header. The header describes the status of toponymic research for the file as a whole. It is a 256-byte character record that includes:

- analyst(s),
- file creation date, and
- comments.

---

* At this stage, area feature centroids and the linear feature endpoints are sufficient. If the file is compiled for map production, detailed coordinates are added later from other DMA data sources.

Comments supplied in the header refer to the stage of research:

- fully researched and ready for merging with the GNDB,
- entirely unprocessed, or
- partially processed, problems isolated to (problems).

8.1.2.b Contents. Contents are in Standard Data Transfer Format (Table 8-1). Between Standard Data Transfer Records is a 256-byte comment field for transmitting relevant information. Examples of comments follow.

- "Other names on this source were not reliable."
- "Appears to be a variant spelling for (name)."
- "Other sources consulted: (list)."
- "This name is correct; the one in the data base is no longer used. (Toponymist's initials included.)"

Table 8-1. Standard Data Transfer Record.

Entity Name	Size (Bytes)
Data Source Name	10 (1)
Number of Characters in Geoname	1
Number of Characters in Non-Anglicized Name	1
Number of Characters in Alias	1
Number of Characters in Province Name	1
Number of Characters in Country Name	1
Names (geoname, non-Anglicized name, alias, province name, country name)	140 (2)
Type of Romanization	1
Date of Data Source	3 (3)
Date of Data Capture	3
Date of Last Update	3
Position	6 (4)
Positional Accuracy	2
Feature Designator	6
Attribute	6
Administrative Code	1
Area Code	1
UTM grid	8
Selected Map Sheet	7
Approved or not Approved	1
Bounding Rectangle	13 (5)
Pointer to File Containing Feature Coordinates	8
Unused	32
	256

(1) The GNDB maintains a dictionary of legal data sources.

(2) If more than 140 characters are required, the next record is an overflow record. All names are stored in this field to substitute one large field with overflow allowances for potentially five large fields with possible overflows.

(3) Dates are numeric strings: ddmmyy.

(4) Position as currently planned is a point (the location of a point feature, the mouth of a river, or the centroid of an area feature) given as two signed numeric strings: +/-dddmmss and +/-ddmmss. Negative indicates latitude South or longitude East, positive indicates latitude North or longitude West.

(5) The bounding rectangle is high and low latitudes and longitudes, with an additional byte indicating if the bounding rectangle is incomplete due to the feature leaving the map.

8.2 Product Data Sets

Product data sets are for generating a geonames product. The only ASP product data set is a Gazetteer File, described here.

8.2.1 Gazetteer File

A Gazetteer File is a fully edited set of geonames for a single country or a geographic region, a digital version of a hardcopy gazetteer. It does not necessarily include all of the region's stored names. When completed, the file is phototypeset and archived.

8.2.1.a Header. The header shows the production history of the gazetteer in progress. The header also contains the explanatory information shown in the Foreword of a gazetteer. Production history includes:

- initial compilation date,
- completion date (if completed),
- responsible party, and
- contributing parties.

Explanatory information includes:

- introduction,
- explanation of feature designation,
- glossary of generic terms,
- information for report errors, and
- transliteration system, if applicable.

8.2.1.b Contents. Contents are alphabetized geonames with:

- feature designator,
- latitude and longitude to the nearest minute,
- UTM coordinates to 8 characters,
- map sheet code, if required, and
- area code.

8.3 Lookup Tables

Lookup tables are read-only files used during processing.

8.3.1 Telegrapher's Code

The Telegrapher's Code reverses transliterations of ideograms.

8.3.2 Dictionary

The Dictionary holds feature designation definitions and definitions of generic terms. It is an on-line toponymic reference.

8.3.3 Transliteration Rules

Transliteration rules are provided for each of the major linguistic groups. They are comprised of the transliteration information shown in a gazetteer of that linguistic area. This file is an on-line toponymic reference and a source of information for compiling Gazetteer File forewords. "Scratchpads"

will be provided for toponymists to add notes, exceptions, and special rules that are not shown in the final published gazetteer.

8.3.4 Coordinate Limits

Coordinate limits are used to detect positioning errors to ensure that a name's coordinates are within its geopolitical boundaries. The data structure is a 180 X 360 matrix, each matrix element corresponding to a 1^0 square on the earth's surface. If the 1^0 square contains only one geopolitical entity, the element indicates which entity (''noncountry'' is a geopolitical entity defining oceans). If the 1^0 square includes more than one geopolitical entity (e.g., includes a piece of an international border), the element would indicate which geopolitical entities are involved. More than two geopolitical entities might lie within a 1^0 square.

8.3.5 Area Code Boundary File

Because area codes used in gazetteers are not generated from a mathematical formula, a file is required to convert latitude/longitude pairs into area codes. There are several data structure alternatives. One approach is to represent area boundaries as lines represented by point sequences. A second approach is to use a variable resolution grid (e.g., use quad trees or other hierarchical grid encodement) with higher resolution for boundary grid cells. This file should be able to determine area codes from geographic coordinates but should not have a level of detail that makes it a small feature file.

8.4 System Accounting

8.4.1 Product Log File

The Product Log File is used by the production supervisor to track the progress of a product through the system and to generate production statistics. The header records of each processing and product data set are stored in the *Product Log File.*

8.4.2 System Log File

The System Log File provides daily, weekly, and monthly statistics of system use, including the following entries:

- process initiation,
- process termination,
- printer initiation,
- printer termination,
- tape initiation,
- tape termination,
- disk initiation,
- disk termination,
- edit station initiation,
- edit station termination, and
- system supervisor console.

For each entry the following information is required:

- entry type,
- user ID,
- data set ID,
- start/stop/time/date,
- device ID/process step ID,
- CPU time,
- file name,

- normal/abnormal termination indicator,
- process ID,
- tape/disk ID number, and
- system supervisor message (where applicable).

8.4.3 Archives Catalog

All completed gazetteers are listed by header record (excluding the Gazetteer Foreword information) for a complete digital product catalog. The Archives Catalog is queried by country and by date.

9.0 INTERFACES TO OTHER SUBSYSTEMS

The Geonames Processing System is a network of processors with special-purpose peripherals. The system network must have either hardware connections or standard data interfaces. Functional and data interfaces are illustrated in Figure 9-1.

9.1 Geonames Data Base/ASP

ASP has the only read/write access to the GNDB. Controlling interaction serves three purposes.

- DMA does not wish to maintain dual GNDBs for each production center. Limited GNDB interaction encourages development of strong batch utilities that facilitate data transfers to DMAAC, to occur through communications link or tape transfer.
- Minimizing interactive activities enhances responsiveness.
- GNDB integrity must be maintained. Therefore, access is limited to the ASP subsystem.

As the only subsystem with read/write access to the GNDB, ASP's chief responsibilities are file compilation and manipulation, and data base update and maintenance.

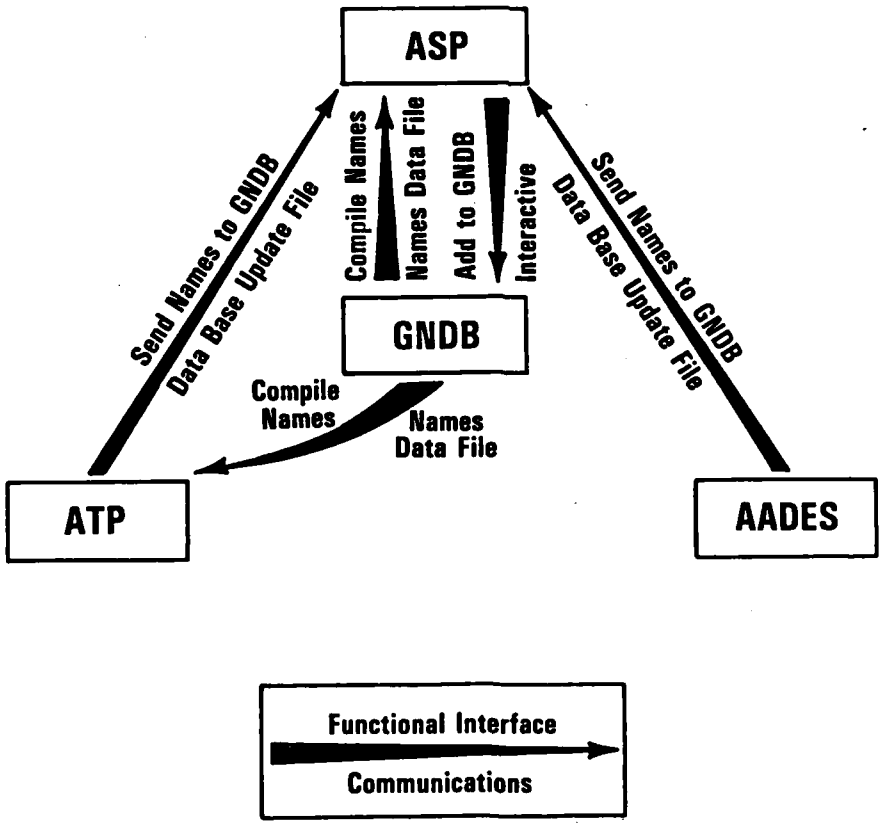


Figure 9-1. Interfaces between the subsystems

Data moves from the GNDB to ASP via batch-compiled disk or tape files, or by interactive command. Data moves to the GNDB from ASP via Data Base Update Files that have been interactively corrected by a toponymist at the ASP workstation.

GNDB query and manipulation functions are used by both ASP and the GNDB. These functions are likely to be part of the GNDB data base management system software. A link is required between ASP's Data Base Update function and the GNDB's data entry function.

9.2 Advanced Type Placement/ASP

ATP and ASP share hardware, software, and files. Map names are originally compiled from the GNDB into a Names Data File using an ASP workstation. Once correctness is assured, a Map Data File is generated from Names Data File information. Corrections to names made during ATP are stored in a Data Base Update file and returned to ASP for supervised entry to the GNDB. Editing software and toponymic utilities are shared by ASP and ATP.

9.3 Automated Alphanumeric Data Entry System/ASP

AADES is a high-volume names data capture system. The data collected by AADES is stored in a Data Base Update File and passes through ASP for supervised entry to the GNDB.

10.0 ASSUMPTIONS AND CONSTRAINTS

ASP operation depends on hardware interfaces and a well-managed GNDB with defined coverage. Details are provided here.

10.1 Hardware Interface Between Subsystems

ASP is a direct-access GNDB terminal. Word processing is performed locally after downloading files from the data base. This functional design assumes no particular hardware manufacturer requirement.

Although a certain amount of tape handling is inevitable in data processing applications, a more direct means of data transfer is preferred. Modem connections are to be used only when distance is a factor, since they tend to be slow, costly, and occasionally inaccurate.

10.2 Level of GNDB Coverage

Names data capture will occur over a period of 10 or more years, due to the number of names to be input. After full coverage is reached, names will change, data sources will improve, and new areas will become of special interest, making continuing maintenance crucial.

Initial data capture must be structured so coverage is gained most effectively. This FDS assumes that in the early years of the Geonames Processing System, product generation will begin with toponymic research, verification, and GNDB updates. If, however, lack of current GNDB coverage causes continued manual names compilation, users will become disillusioned by the inconvenience incurred by a system that is intended to reduce their workload.

A DMA commitment to full GNDB loading within a limited timeframe is required for system success. With a concerted initial effort at data capture, the GNDB can be an important boost to geonames processing productivity at DMA.

10.3 Non-Roman Script

A capability to process ideographs is not recommended for DMA's limited ideograph requirement. All such methods reviewed in this study require specially trained personnel or specialized equipment. Under the current development plan, ideographs can be stored as images (bit maps) within names records. Thus, while a name cannot be accessed or sorted by ideograph, the ideograph can be accessed by its Romanized version for toponymic use or reproduction.

11.0 PERFORMANCE REQUIREMENTS

ASP performance requirements state the level of increased operational capability over current capability with introduction of the subsystem.

11.1 File Management

File Management must be capable of:

- 10 data set creations per hour,
- 20 data set retrievals per hour, and
- 10 data set deletions per hour

on data sets with an average size of 640KB. Data set size ranges from less than 280KB to 1MB. File Management allows 100 data sets to be stored at once, and 10,000 data sets to be maintained at once (some file management may be performed by the GNDB). File Management must provide a minimum of 10MB on-line storage for each workstation.

11.2 Job Management

Job Management must accommodate records on 1000 data sets in on-line storage. Job Management must allocate work so queued tasks are sent to the first available processor that can perform the work. System components requiring human interaction may remain idle due to personnel shortage.

11.3 ASP Applications

ASP linked to the GNDB introduces automation to previously manual tasks. There must be an overall 30% improvement in timing with ASP over the current corresponding manual methods.

Timing improvements cannot be accurately stated, since the bulk of the work involved is research and decision-making by analysts. The expected 1989 workload will not be forecast in this report. The ASP subsystem will support 20 workstations operating concurrently without a noticeable increase in response time. It is estimated that of these stations, three will be involved in major revisions and the rest in low-volume queries.

11.4 Operational Requirements

All ASP functions are called with a minimum of keystrokes. Frequently used command formats are stored in executable files so highly repetitive work is avoided.

ASP must adjust to user expertise by providing menus and help sequences for beginners and a terser dialog for experts.

ASP must allow interactive edit and viewing of diacritics, multiple fonts, special symbols, and kerning.

REDS must be easily used and recalled. One of the following means is suggested:

- color-coded keyboard templates,
- interchangeable keyboards, or
- a small CRT that displays the current REDS.

12.0 HARDWARE REQUIREMENTS

ASP requires a computer, two tape drives, on-line mass storage, a console for operators and programmers, workstations, and graphics printers. The system hardware must handle the processing requirements presented in Sections 7 and 11 by carrying out the functions described in Sections 2 through 6 within the assumptions and constraints detailed in Section 10. ASP must be modularly expandable.

12.1 Computer

A computer is required for each production center. The computer need not necessarily be dedicated to ASP if workstations have local processing power. Computer memory must be capable of handling the concurrent production load specified in Section 11. It must also be capable of expanding to handle twice the specified production load. Spare unused capacity must be provided for program development. Twenty-five percent of the memory delivered must exceed the amount required by ASP applications and system demands.

12.2 Tape Drives

ASP needs two 9-track, 1600/6250 bpi tape drives for each production center.

12.3 On-Line Mass Storage

On-line mass storage must meet the requirements of file management presented in Section 6. Twenty-five percent of the mass storage delivered must be available to support program development.

12.4 Operator and Programmer Consoles

Two or more consoles (at minimum, one for each production center) are required for computer operators who mount and assign tapes. The consoles are the operators' interface with Job Management and File Management. There must be at least one alphanumeric CRT and keyboard for software maintenance and program development.

12.5 Workstations

Each workstation consists of:

- a terminal with display controls and programmable function keys to support the editing software, and
- a graphics printer.

12.5.1 Terminals

The workstation terminal must be able to display international geonames in their final form, complete with diacritics and kerning. It must aid the analyst's recall of function key configuration. The terminal must provide:

- readable proof copy at a rate of no less than 30 cps,
- nonglare terminal screen for display,
- an echo capability between keyboard and display,
- a minimum display speed of 120 cps,
- minimum 80-character display lines,

- display of diacritics and special symbols,
- display of the full ASCII 96-character set, and
- transmittal of the 128-character ASCII set.

12.5.2 Printer

A printer is required for proof copies. It must print diacritics and special characters and have minimum 80-character lines. A nonimpact printer/plotter is desirable for this application due to speed, quietness, and versatility of format.

12.6 Plotter

A graphics hardcopy device is desirable for proof copies of graphics. A pen plotter or a combination printer/plotter would answer this need.

12.7 Federal Information Processing Standards

As required by Federal Procurement Regulation (FPR) 1-4.1108.5 and Federal Property Management Regulations (FPMRs) 101-36.1304 and 101-36.1305, all equipment specified must conform to the following Federal Information Processing Standards (FIPS).

12.7.1 Interchange Codes and Media

FIPS PUB 1—Code for Information Interchange.

The system, upon receiving a hardware or software command, must accept data on magnetic tape, paper tape, or any other input media covered by an approved FIPS PUB in ASCII code, and the collating sequence prescribed in FIPS PUB 1, and in the format prescribed in FIPS PUBS 2,3 or other applicable FIPS PUBS. Such data may be translated, if necessary, into a form which the proposed equipment can internally process provided that, upon receiving a hardware or software command, the output of the processed data to magnetic tape, paper tape, and other output media will be in the ASCII code and collating sequence prescribed in FIPS PUB 1 and in the format prescribed in FIPS PUBS 2,3, or other applicable FIPS PUBs.

FIPS PUB 15—Subsets of the Standard Code for Information Interchange.

Printers; display devices; data acquisition, preparation, and transcription devices; data communication terminal devices; and other data processing or communications equipment not requiring the full 128-character set of the Federal Code for Information Interchange, FIPS PUB 1, must conform to one of the approved character Subsets of the Standard Code for Information Interchange, FIPS PUB 15. Printers of the "chair" or "train" or other replaceable symbol technology may also be provided with optional subsets having a different number of characters from those specified in FIPS PUB 15 in order to increase the printer's speed as required for local use, provided the ability to interchange information by the selected character subset (FIPS PUB 15) is retained in the data processing system.

FIPS PUB 25—Recorded Magnetic Tape for Information Interchange (1600 CPI, Phase Encoded).

All 9-track digital magnetic tape recording and reproducing equipment employing ½-inch wide tape at the recording density of 1600 CPI (phase encoded), including associated programs, shall provide the capability to accept and generate recorded tapes in compliance with the requirements set forth in FIPS PUB 25.

FIPS PUB 35—Code Extension Techniques in 7 or 8 Bits.

All coded character sets which require control function and/or graphic symbols that are not included in the 128 characters of ASCII will be implemented through the use of the code extension methods and techniques as described in FIPS PUB 35.

FIPS PUB 36—Graphic Representation of the Control Characters of ASCII (FIPS PUB 1).

All applicable equipment that prints or displays graphic representations of any or all of the control characters of ASCII (FIPS PUB 1) or of the characters "space" or "delete" must comply with the requirements set forth in FIPS PUB 36. This standard also applies to equipment that prints these graphic representations on media such as perforated tape, punched cards, or listing.

FIPS PUB 50—Recorded Magnetic Tape for Information Interchange 6250 CPI (246 CPMM), Group Coded Recording.

All applicable digital magnetic tape recording and reproducing equipment which employs ½-inch-wide (12.7 mm) magnetic computer tape at the recording density of 6250 characters per inch (246 characters per millimeter) group-coded recording, including associated programs, shall provide the capability to accept and generate recorded tape in compliance with the requirements set forth in FIPS PUB 50.

12.7.2 Transmission

FIPS PUB 16-1 (FED-STD 1010)—Bit Sequencing of the Code for Information Interchange in Serial-By-Bit Data Transmission.

All applicable equipment or services transmitting in a serial-by-bit, serial-by-character mode must be capable of bit sequencing as prescribed in FIPS PUB 16-1/FED-STD 1010 for the transmission of the Standard Code for Information Interchange, FIPS PUB 1, at the interface between data terminal equipment and data communication equipment.

FIPS PUB 17-1 (FED-STD 1011)—Character Structure and Character Parity Sense for Serial-By-Bit Data Communication in the Code for Information Interchange.

All applicable equipment transmitting in a serial-by-bit, serial-by-character synchronous or asynchronous mode must be capable of transmitting the character structure and sense of character parity prescribed in FIPS PUB 17-1/FED-STD 1011 for the transmission of the Standard Code for Information Interchange, FIPS PUB 1, at the interface between data terminal equipment and data communication equipment.

FIPS PUB 18-1 (FED-STD 1012)—Character Structure and Character Parity Sense for Parallel-By-Bit Data Communication in the Code for Information Interchange.

All applicable equipment transmitting in a parallel-by-bit mode must be capable of transmitting the character structure and sense of character parity prescribed in FIPS PUB 18-1/FED-STD 1012 for the transmission of the Standard Code for Information Interchange, FIPS PUB 1, at the interface between data terminal equipment and data communication equipment.

FIPS PUB 22-1 (FED-STD 1013)—Synchronous Signaling Rates Between Data Terminal and Data Communication Equipment.

All equipment or services that are employed in conjunction with synchronous data communication equipment designed to operate on binary encoded information in either serial or parallel fashion over voice grade communication channels of nominal 4-kHz bandwidth must comply with FIPS PUB 22-1/FED-STD 1013.

FIPS PUB 37 (FED-STD 1001)—Synchronous High Speed Data Signaling Rates Between Data Terminal Equipment and Data Communications Equipment.

All equipment or services that are employed with synchronous data communication equipment designed to operate on binary coded information over wide band communication channels must comply with FIPS PUB 37/FED-STD 1001.

APPENDIX A

BIBLIOGRAPHY

Alden, David G.; Daniels, Richard W.; and Kanarick, Arnold F. "Keyboard Design and Operation." *Human Factors* 14 (1972): 275-293.

Augustine, Robert F.; Caldwell, D.R.; and Strife, D.E. "A Prototype Geographic Names Input Station for the Defense Mapping Agency." ***Auto Carto V***, August 1982.

Bie, Stein W. "Organizational Needs for Technological Advancement: The Parenthood of Autocartography Revisited." ***Auto Carto Six***, October 1983.

Bohme, Rolf. "An Integrated System of Processing Geographical Names." *International Yearbook of Cartography* 13 (1973): 97-109.

Boyle, A. Raymond. "Report on Symbol and Name Manipulations and Placement." Graphic Systems Design and Application Group, Electrical Engineering Department, University of Saskatchewan. August 1974.

Carroll, John M. "Presentation Form in User-Interface Architecture." *Byte* 8 (Dec. 1983): 113-119.

"Development of an Automated Cartographic Capability," Washington, D.C.: DMAHTC, April 1982.

"The Feasibility of Establishing an Automated Chart Production Process." Washington, D.C.: DMAHTC, October 1982.

Hayes, P.; Ball, E.; Reddy, R. "Breaking the Man-Machine Communication Barrier." *Computer* 14 (March 1981): 19-30.

Hershey, A. V. "Calligraphy for Computers." Dahlgren, Virginia: Naval Weapons Laboratory Report #2101, 1 Aug 1967.

Hershey, A. V. "FORTRAN IV Program for Cartography and Typography." Dahlgren, Virginia: U.S. Naval Weapons Laboratory, 1969.

Hershey, A. V. "A Computer System for Scientific Typography." *Computer Graphics and Image Processing* 1 (December 1972): 373-385.

Jablinski, R.; Strife, D.; Gaar, K.; and Moore, J. "Names Type File System." Consulting Report for the USAETL Project #P0013. April 1983.

Opalski, William E. "Automation of Foreign Place Names at the United States Defense Mapping Agency." Unpublished Report, USAETL, Ft. Belvoir, Virginia: 1980.

Payne, Roger L. "Geographic Names Information System." U.S. Geological Survey Circular 895-F, 1983.

U.S. Army Engineering Topographic Laboratories. "Error Detection in a Toponymic Database: a Case Study Using a Database Management System." Ft. Belvoir, Virginia. Unpublished Report, 1980.

Wolcott, H. M. And Hilsenrath, J. (1976) "A Contribution to Computer Typesetting Techniques: Tables of Coordinates for Hershey's Repertory of Occidental Type Fonts and Graphic Symbols." National Bureau of Standards Special Pub. #424, U. S. Department of Standards, Washington, D.C.

APPENDIX B

ORIGINAL DMA REQUIREMENT STATEMENTS

DMA's four original requirement statements date from 1979. The principal portions of these statements are reproduced in this appendix. The first three requirements are from DMAHTC, the last is from DMAAC.

B.1 Requirement 1—Automated Alphanumeric Data Entry

B.1.1 Background

B.1.1.a Operational Setting

The present capability for inputting text and numerical data into a computer for processing consists of keypunching cards, keying a scope terminal, or using a highly constrained Optical Character Recognition device. A significant amount of resources are expended for data preparation of such items as Time and Attendance Records, Topo Data Library System, Bathymetric Data Library System, Geographic Names, Positional Data, Imagery Data Files, and Hydrography Files.

B.1.1.b Deficiency

This labor-intensive method is error-prone and results in many hours being spent in correcting data that has been inputted. All of the data that was originally scheduled to be in the Topo Data Library System (TDLS) has not been inputted due to the lack of resources (there are approximately 600,000 unique items in TDLS). There are numerous other databases in the planning stage, and an effective, economical way must be found to input data or the systems will never be fully developed.

B.1.1.c Related Work

R&D Project "Voice Recognition." RADC has been working on a system to automatically enter digital sounding data onto a magnetic tape. This system is activated and operated by an operator who speaks into a microphone. DMAHTC is procuring a "Data Entry Edit System" to support several existing and proposed data base activities. This is to be a minicomputer system with interactive entry/edit terminals; however, it still must be manually fed.

B.1.2 R&D Objectives

Investigate and analyze current computer I/O devices and requirements at DMAHTC and develop a cost-effective, 99% error-free system to convert alphanumeric data into computer-readable form.

B.2 Requirement 2—Geographic Names Database System

B.2.1 Background

B.2.1.a Operational Setting

There are two branches in the Geographic Names Data Division, Scientific Data Department.

- The Toponymic Branch performs geographic and toponymic research that leads to the development of policies, procedures, and directives for the treatment of geographic names. They maintain a worldwide Geographic Names Data Base on 4-1/2 million index cards and the

DoD Foreign Place Names File. They are responsible for the publication of gazetteers, politico/administrative studies, and glossaries. Approximately 270 country gazetteers have been published to date. Gazetteer information is stored on magnetic tapes.

- The Applications Branch assembles geographic names data and related descriptive information for DMA topographic maps, nautical charts, placename indexes, and special purpose products. They maintain a population file, a boundary inventory, and provide identification, description, and designation categories of all natural and cultural features requiring labels on DMA graphic products. They also develop specifications for names presentation on DMA cartographic products. Names placement data is furnished to the Symbols and Names Placement System (SNAPS) on handwritten data sheets. SNAPS personnel convert the data to magnetic or paper tape for names placement.

B.2.1.b Deficiency

The development of a digital Geographic Names Database into a common, efficient, and reusable format is required to minimize rework of the same data by the two branches. Significant savings in time and production funds may be realized through the design and implementation of this data base. DMA's 270 gazetteers are to be updated at the rate of 6–10 per year. Some gazetteers are 10–25 years old, and their names cannot be used for names placement. This is due to the limited amount of resources available. Modern, locally used names can be found by researching text, magazines, phone books, foreign-published maps, or other sources. The result is that some names appearing on DMA graphic products are not the same as those in the gazetteers. The current magnetic tape method of producing a gazetteer has no provisions for furnishing diacritics, special alphanumeric characters, and lower case letters. The tape used to drive SNAPS is normally destroyed after use, since it is formatted to place names on a particular map sheet at a prescribed scale. This tape is not reusable if a map of a different scale is made of the same area. The magnetic tapes used for gazetteer publication cannot be used for names placement because they do not provide geographic coordinates to arc seconds, population data, type font size, style, color, and placement instructions. The country gazetteers are printed in upper-case letters, while names on maps are in upper and lower case.

B.2.2 R&D Objective

R&D's goal is to develop a Geographic Names Database System that will be economically responsive to the needs and requirements of both Branches in the Geographic Names Database Division. Possible candidates for elements of system hardware that could be used are the Electron Beam Recorder, the Cathode Ray Tube Print Head, and ETL's interactive 3-D view graphic system. The applications software and file structure must be designed to enable rapid update and quick response to queries.

B.3 Requirement 3—Advanced Symbol Processing System

B.3.1 Background

B.3.1.a Operational Setting

The type placement and symbol processing situation at DMAHTC is presently in a chaotic state. The present Photon type placement systems are outdated and replacement parts are nonexistent. The Geographic Names Files currently consist of trig lists, card files, catalogs, and several other archaic methods of retaining information. Presently, diacritics and special notations are corrected or initially performed by manual methods, with no records kept on what is done. Digital data are used more and more, but the present fonts, type, and words are not in the proper format or digitized to be used in these newer data types.

B.3.1.b Deficiency

Commercially available or R&D-developed type placement systems do not satisfy DMAHTC requirements for diacritics and kerning. For effective productivity, operators must view the diacritics as they will appear in final type. There is need within DMAHTC to establish a library or disk file of the most commonly used fonts, words, and type. To fully utilize the CRT Print Head System, the Electron Beam Recorder, or future systems, this library or disk file is a must. In the near future, several systems will need to draw from standardized digitized files. An interface between the old system SNAPS (Symbols and Names Placement System) and the newer systems being delivered is lacking.

B.3.1.c Related Work

R&D—Type Composition Console
R&D—Geo (Geographic) Names Files
R&D—Font Digitization
R&D—SNAPS to CRT Print Head Conversion
TIP—OT&E of the CRT Print Head System

The Type Composition Console is an R&D item or system being developed for DMAAC. Presently undergoing Operational Testing and Evaluation (OT&E) at DMAAC, it lacks a diacritics or kerning capability. Geonames Files are an R&D item to develop a capability for storing approximately 10,000 commonly used names on disk or magnetic tape, with easy access and retrieval functions. Font digitization, a USAETL in-house effort, is digitizing the most commonly used fonts and type within DMA, again with an easy retrieval capability. SNAPS to CRT Print Head conversion is being accomplished under R&D, but it should be expanded to cover other type or word placement systems, as well as the Electron Beam Recorder. TIP is testing and evaluating the CRT Print Head newly installed on the Gerber Precision Plotter and Concord Plotter at DMAHTC.

B.3.2 R&D Objective

To provide DMAHTC with a General Purpose Symbol Processing System suitable for use with a variety of functional operations (i.e., Names and Placement System, Notice to Mariners System, Geographic Names System, etc.). This system can provide the basis for a replacement of the Photon and drum coordinatograph portions of the present Names Placement System, and provide the proper interfaces for easy access to standardized digitized fonts, names, words, and type by new equipment.

B.4 Requirement 4—Digital Type Composition and Placement

B.4.1 Background

B.4.1.a Basic Objective

The basic objective is to develop an all-digital system for the composition and placement of typographical names shown on chart products.

B.4.1.b Requirements

This system is needed to prepare names information depicted on Air Target Material, and Navigation and Planning Chart products used in the operation of U.S.A.F. weapon systems, as well as in support of space exploration programs.

B.4.2 R&D Requirement

Present DMA typographical systems compose and position the characters that comprise geographic names and identifiers via key-board cursor and aperture systems. This requirement is to address development of a more advanced system that would permit these functions to be performed more interactively and efficiently, through increased use of electronic display technology.

APPENDIX C

TOPONYMIC DEFINITIONS*

Geographic entity: a more or less permanent place, area, or feature having a recognized id ·ity and a fixed location on or near the earth's surface. It may be either natural or cultural (ma nade), or it may combine both of these attributes. It may be either tangible or intangible and hav definite, indefinite, or variable intent. Examples: mathematical lines (Tropic of Cancer) and points (North Pole), world zones and belts (Western Hemisphere, Horse Latitudes, Heartland), continents, oceans, ocean currents, regions, countries, administrative areas, populated places, sections of populated places, settlements, localities, missions, factories, schools, buildings, streams, farms, plantations, ranches, seasonal dwellings, archaeological and other sites, monuments, landings, canals, ditches, streets, fords, bays, roadsteads, channels, shoals, glaciers, mountains, plains, islands, caves, cliffs, escarpments, nunataks, seamounts, volcanoes, forests, basins, mines, deltas, peninsulas, capes, and many others.

Geographic name: a proper noun used to identify a geographic entity. The name may consist of a specific element only (Skagerrak) or it may also contain a generic term and other words (Lake of the Woods). In many cases a geographic entity is correctly identified by two or more names; for example, by a local official name (Roma) and a conventional name (Rome) or by long and short forms, both equally correct (Danube and Danube River, Tuxpan and Tuxpan de Rodriguez Cano, Bedzin, and Powiat Bedzinski). Many foreign names contain essential modified letters, diacritical marks, and punctuation, without which they would be incorrectly spelled. Geographic names fall into several categories and subcategories, namely: BGN-approved name, alternate name, variant name, or other accepted name.

BGN-approved name: a geographic name approved by the Board on Geographic Names, either individually as a decision or en masse with others as an official or provisional standard name. Approved names are of the following types: conventional names, local official names, and anglicized names.

Conventional name: a name widely used by English speakers and applied to a well-known geographic entity. The entity may be either international or entirely within a single foreign country or dependency. If the latter, the conventional name is not identical in form and spelling with the local official name. Examples: Atlantic Ocean, Lisbon, Kingdom of Norway, Danube River.

Local official name: (a) a Roman-alphabet name as used officially to identify a geographic entity in a Roman-alphabet area (Oslo), or (b) a Roman-alphabet name derived by the application of a BGN Romanization system to the name of an entity as officially written in the language of a non-Roman-alphabet area (Moskva).

Anglicized name: (a) an English language name lacking the usage and recognition of a conventional name, which is applied to an entity in an area over which the U.S. Government recognizes no sovereignty; e.g., Antarctica (Bjelland Point) and the floor of the ocean (Kushiro Canyon), or (b) an English-language name, also lacking the usage and recognition of a conventional name, which is applied to an entity in a non-English area. Example: Tung-hsing Multinational Autonomous Region (China).

Other acceptable name: a geographic name that has been obtained from authoritative sources and by the application of BGN policies, including use of the appropriate BGN-approved Romanization system, but which has not been approved as a decision or standard name by the BGN. These names are of the same types and have the same attributes as the BGN-approved names described under above.

---

* From draft report of the "Study on the Automation of Geographic Names," Defense Mapping Agency, 24 October 1968.

Alternate name: For a limited number of geographic entities, two or more BGN-approved or other acceptable names are equally approved and acceptable for primary use on maps, charts and other materials; e.g., Roma, Rome, Passo del San Gottardo and Saint Gotthard Pass, Oran and Region d'Oran. Such names are called alternate names. Each alternate must be identified by language, area, or other label so the user may make the proper choice of alternatives; e.g., Spanish, Lapse, conventional, Chinese, Argentinian, long form, etc.

Variant name: a name by which an entity is or has been known other than a BGN-approved or other acceptable name. Variant names include former names, spellings arrived at by use of an unapproved Romanization system, names or spellings reflecting a sovereignty not recognized, or language not authorized for the area in which the entity is located, misapplied names, typographical errors, and any other type that deviates from the approved name(s) and spelling(s). Variant names are not approved for primary use, although they may be used in secondary or parenthetical position to aid in the identification of an entity; e.g., on a map it may be desirable to provide the Turkish name of the Aegean Sea or the Biblical name of a village in Israel as matter of additional useful information.

Cross-reference: a variant or alternate name entry in a name file or listing, which makes reference to a BGN-approved or other acceptable name for the same identical geographic entity by use of the word "see"; e.g., Fegboho see Fegbobo, populated place, 9°14'N, 5°40'W, Ivory Coast. In the case of the alternate names the reference is to the approved name listed first in the approved-name entry; e.g., Rome see Roma, populated place, 41°54'N, 12°29'E, Italy. Cross-references contain the same identifying information as approved-name entries. Ordinarily, to avoid cluttering files, lists, and other materials with unnecessary entries, variants that differ from approved names and other variants only in diacritical marks, word spacing, capitalization, punctuation, and common generic terms not used in cross referencing, or otherwise identified or listed.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS None
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NORDA Report 101	5. MONITORING ORGANIZATION REPORT NUMBER(S) NORDA Report 101
6. NAME OF PERFORMING ORGANIZATION Naval Ocean Research and Development Activity	7a. NAME OF MONITORING ORGANIZATION Naval Ocean Research and Development Activity
6c. ADDRESS (City, State, and ZIP Code) Ocean Science Directorate NSTL, Mississippi 39529-5004	7b. ADDRESS (City, State, and ZIP Code) Ocean Science Directorate NSTL, Mississippi 39529-5004
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Defense Mapping Agency 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) HQ/STT Washington DC 20305	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. 64710B; PROJECT NO.; TASK NO.; WORK UNIT NO.
11. TITLE (Include Security Classification) The Geonames Processing System Functional Design Specification, Volume 4: Advanced Symbol Processing	
12. PERSONAL AUTHOR(S) Gail Langran	
13a. TYPE OF REPORT Final	13b. TIME COVERED From _____ To _____
14. DATE OF REPORT (Yr., Mo., Day) March 1985	15. PAGE COUNT 45
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Maps, computers, software systems

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report describes the Geonames Processing System attributes and serves as a basis for understanding between the user and the developer. The subsystems referred to are: Advanced Symbol Processing, Advanced Type Placement, Geographic Names Data Base, and Automated Alphanumeric Data Entry System.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ■ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Gail Langran	22b. TELEPHONE NUMBER (Include Area Code) (601) 688-4449 22c. OFFICE SYMBOL Code 351

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

END

FILMED

1-86

DTIC